DOCUMENT RESUME

ED 253 837 CG 018 040

AUTHOR Moberg, D. Paul
TITLE Evaluation of Prevention Programs: A Basic Guide for Practitioners.
INSTITUTION Wisconsin Clearinghouse, Madison.
SPONS AGENCY Wisconsin State Dept. of Health and Social Services, Madison.
PUB DATE 84
NOTE 33p.; Sponsored and funded by the Prevention Committee of the Division of Communnty Services.
AVAILABLE FROM Wisconsin Clearinghouse, 1954 East Washington Avenue, Madison, WI 53704.
PUB TYPE Guides - Non-Classroom Use (055)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS Data Collection; *Evaluation Methods; Evaluation Utilization; *Formative Evaluation; Guidelines; *Prevention; Program Effectiveness; *Program Evaluation; *Summative Evaluation

ABSTRACT

This guide is intended for professionals, laypersons, funding agents and others involved in planning and delivering local prevention services. Chapter 1 defines prevention, and differentiates between prevention strategies and programs targeted toward individuals or to general populations. Program evaluation and evaluation research are defined and basic concepts of program evaluation are identified. In chapter 2 specific uses of prevention program evalation are described and related to the concerns of those requesting the evaluation. Chapter 3 outlines nine recommended steps for planning and carrying out any program evaluation. Chapter 4 discusses ongoing monitoring systems, consumer satisfaction surveys, implementation analysis, and participant observation and qualitative interviewing as methods of program monitoring and process evaluation. Chapter 5 discusses uses of experimental control studies, uncontrolled follow-up studies, and pre- and post-test design strategies in outcome evaluation. Chapter 6 provides a summary including precautions for using program evaluation. A glossary of terms, a data collection instrument checklist, and sample monitoring forms are appended. (LLL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED253837

EVALUATION OF PREVENTION PROGRAMS: *A Basic Guide for Practitioners*

by D. Paul Moberg

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

A Friedman

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Sponsored and Funded By:
The Prevention Committee of the
Division of Community Services
Wisconsin Department of Health and Social Services

CG 016040

Acknowledgments

This document was drafted by Paul Moberg under contract to the DCS Prevention Committee. Editing was done by Richard Yoast and Lynn Tolcott of the Wisconsin Clearinghouse; graphic design and production was done by Jane Tenenbaum. A subcommittee of the entire prevention committee, consisting of Joseph Stafford, Sharon Zahradka, Sue Robbins, James D. Kloss, Catherine Henry, and Mary Conroy was extensively involved at all points in seeing this project through to completion. Many of the ideas presented in this guide were developed in interaction with Michael Klitzner, who also provided suggestions based on an earlier draft.

Note to Reader: The first chapter of this guide presents definitions of prevention, program evaluation, and program evaluation concepts; Chapter 2 explains the uses of evaluation. Readers wanting to get right to the "how to's" of evaluation are advised to go directly to Chapter 3.

© 1984 Board of Regents of the University of Wisconsin System for the Wisconsin Clearinghouse

Edited and Published by the Wisconsin Clearinghouse, 1954 East Washington Avenue, Madison, Wisconsin 53704-5291, 608/263-2797.

EVALUATION OF PREVENTION PROGRAMS: *A Basic Guide for Practitioners*

Contents

CHAPTER 1: INTRODUCTION, 1
Prevention Defined, 1
Program Evaluation Defined, 2
Basic Program Evaluation Concepts, 3

CHAPTER 2: WHY EVALUATE PREVENTION PROGRAMS? 5
Uses of Evaluation, 5

CHAPTER 3: GETTING STARTED: PLANNING PROGRAM EVALUATION, 7
Step 1: Identify and Organize Relevant Information Users and Decision-Makers, 7
Step 2: Identify and Refine the Relevant Evaluation Questions, 8
Step 3: Specify Program Goals and Objectives, 8
Step 4: Select Evaluation Method Appropriate to the Evaluation Questions, 9
Step 5: Develop an Evaluation Plan, 11
Step 6: Pilot Test the Evaluation System, 12
Step 7: Implement the Evaluation Plan, 12
Step 8: Summarize, Analyze and Interpret the Data, 12
Step 9: Disseminate and Use the Findings, 15

CHAPTER 4: PROGRAM MONITORING AND PROCESS EVALUATION: A COMPREHENSIVE VIEW, 16
Ongoing Monitoring Systems, 16
Consumer Satisfaction Surveys, 17
Implementation Analysis, 17
Participant Observation and Interviewing, 17

CHAPTER 5: OUTCOME EVALUATION, 19
Experimental Control, 19
Uncontrolled Follow-up Studies, 20
Immediate Outcomes, 20

CHAPTER 6: CAUTIOUS EXPECTATIONS FROM EVALUATION, 21

Reference Notes, 23
Appendix A: Glossary of Terms, 24
Appendix B: Data Collection Instrument Checklist, 26
Appendix C: Sample Monitoring Forms, 26

CHAPTER 1
INTRODUCTION

This guide is for professionals and lay persons directly involved in planning and delivering local prevention services. However, many of the techniques and ideas presented are also useful to funders and others who manage prevention services at any level.

The major purposes of this booklet are:

- To delineate the area of prevention.
- To educate the reader to the full range of evaluation techniques available, while emphasizing those which are relatively simple to implement.
- To provide the reader with a set of realistic expectations regarding the promise and limitation of prevention evaluation.
- To help the reader match evaluation approaches to their program, information needs, and available resources.
- To direct the reader to additional resources, and indicate points at which outside consultation may be needed.

Before proceeding with these tasks, definitions of our subject matter (i.e., prevention, program evaluation) are in order.

Prevention Defined

The term "prevention" is used throughout this guide to distinguish it from intervention, treatment, and rehabilitation programs which have in the past been called "secondary" or "tertiary" prevention. The goal of prevention is to enable individuals to maintain and improve their health and well-being, rather than to intervene after health or well-being has been disrupted (as with intervention and treatment programs). Prevention can occur at the community level as well as at an individual level. It establishes and strengthens community environments which promote health (in its broadest sense) and reduces risks to the individual.

The Wisconsin Department of Health and Social Services (DHSS) provides the following definition:

> *"Prevention policies, programs and activities are those which: 1) enable people to stay healthy; (i.e., maintain and improve physical, emotional, social and intellectual development); and 2) enable communities to strengthen environments which promote health and to change those conditions which predispose individuals to develop problems."*[1]

Similarly, the Wisconsin Prevention Network (WPN) defines prevention as:

> *". . . a process which promotes health by empowering people with resources necessary to confront complex, stressful life conditions and by enabling individuals to lead personally satisfying, enriching lives."*[2]

The major characteristics of prevention, as distinct from other components of human services, are therefore threefold:

- A stress on enabling the individual to maintain or improve his/her level of health, under a broad conception of health, which includes physical, emotional, social, and intellectual functioning; (i.e., full development of human potential).
- A focus on individuals and communities *before* an identified problem occurs.
- Involvement on a community or environmental (as well as individual) level to alter conditions detrimental to health and to strengthen conditions which promote health.

Prevention strategies often include information dissemination, education, provision of positive alternatives, community organization and development, effecting environmental change, and the formation of social policy. At least three of these six strategy areas should be included in any prevention effort to be viable.

Prevention strategies are often *generic* in nature. Since prevention programs often deal with basic human needs rather than specific problem areas, the same program strategy may be useful in preventing a number of categorical problems. For example, alcohol and other drug abuse and juvenile delinquency can be targeted simultaneously by programs such as Positive Youth Development.

1

Programs which specifically target categorical problems also have their place. For example, the prevention of birth defects characterizing Fetal Alcohol Syndrome (FAS) is unlikely to be accomplished by a generic prevention strategy which targets general health enhancement and development of human potential. On the other hand, school based programs which teach middle school students decision-making skills, interpersonal communication, and enhance self-esteem may well play a part in simultaneously preventing alcohol and other drug abuse, vandalism, juvenile delinquency, and many other problem behaviors associated with adolescent development. Some problems (e.g. mental disorder, effects of stress, child abuse and neglect, coronary problems) could be directly and significantly altered by generic prevention programs aimed at health promotion and human development, while others (e.g., lead poisoning, drunk driving) would be more influenced by categorical programs which specifically target one problem.

Prevention programs may target persons specifically at risk (e.g., pregnant women, families with unemployed heads of household, DES daughters) or may target general populations for whom no elevated risk is present. In the first instance, specific efforts would be to prevent the development of the problems for which specific groups are at high risk (e.g., education of pregnant women regarding alcohol use in order to prevent Fetal Alcohol Syndrome; enhancement of family communication skills and stress reduction in families with an unemployed household head to prevent child and spouse abuse and alcohol/drug problems; periodic medical examination of DES daughters to prevent development of cervical cancer). General populations, not at specific risk, are also valid targets of prevention under the assumption that some level of risk is present for all and the view that the true goal of prevention is general human development and well-being, which all can benefit from.

Program Evaluation Defined

Human service practitioners at all levels are frequently involved in evaluative activities. Throughout all phases of program planning, implementation, and service delivery the responsible practitioner asks herself whether or not her program is meeting the needs of its target population, whether program services are being delivered efficiently, and whether the program is accomplishing what it had hoped. The practitioner learns by providing service and through interactions with various constituents (recipients, other professionals, funders, community groups, etc.) how well accepted the program is, what problems there are, where a program's strengths lie, and what alterations need to be made. As a result of this, programs evolve and change over time.

While this evaluative process can lend to many program adjustments, it suffers because it is not systematic in its data collection and analysis. Many issues which could be addressed in formal evaluations are overlooked. The objectivity required in evaluation efforts is unlikely to be adopted and conclusions are easily distorted. For these reasons, **we suggest that the evaluator role be taken on explicitly, and that it be viewed as essential to but separate from ongoing service delivery.**

Making the evaluative process explicit and systematizing data collection and analysis are key elements of the activities known as "program evaluation" or "evaluation research." A general definition of this process is:

> *"Program evaluation is the systematic collection, analysis and interpretation of data for the purpose of determining the value of a social policy or program, to be used in decision-making about the policy or program."*

There are three key aspects to this definition. First, data collection is **systematic.** It is planned for in advance and follows specific procedures for sampling, documentation, verification, and validation. Second, the **value or merit of a program or policy must be determined,** either relative to other alternative activities (or no activity), or to a set of goals or criteria of value or success. Third, the purpose of evaluation sets it aside from other forms of research—**evaluation is to be used in decision-making** about the program. Use in decision-making can be at any level, from program development by direct service practitioners to decisions regarding the establishment of national policy.

While our definition is a general one encompassing all types of formal evaluation, it is useful to further distinguish between "program evaluation" and "evaluation research." **Program evaluation** is *for* the program while **evaluation research** tends to be *of* the program. Program evaluation seeks to generate information which is *useful* to ongoing program planning, development, and administration,[3] while evaluation research seeks to determine the effectiveness of program models and tests theories of prevention. Program evaluation describes program accomplishments, processes, and problems, while evaluation research tends to focus primarily on the potential effects of new policies or programs. Ideally, program evaluation serves as an integral part of rational program management and accountability, reducing uncertainty (but never eliminating it) in decision-making concerning day-to-day

operations and long-term planning. Evaluation research tends to be farther removed from these day-to-day concerns.

For purposes of this guide a sharp distinction between program evaluation and evaluation research has been made. However, in practical applications and in the evaluation literature the distinction becomes blurred. It is perhaps best to visualize a continuum of evaluation, with most evaluation efforts somewhere between the two polar anchors. The bulk of the material in this guide falls on the "program evaluation" side of the continuum, presenting relatively simple techniques for local level internal evaluation of programs. However, a few topics (such as experimental design) are also mentioned to provide a basic understanding of the issues involved in evaluation research applications.

Basic Program Evaluation Concepts

As the national concern with program evaluation has accelerated over the past fifteen years, evaluation models, theories, and approaches have proliferated. The following are some conceptual distinctions relevant to prevention which have arisen from all this.

Summative vs. Formative Evaluation[4]

This is a useful distinction in considering the purpose of an evaluation effort. **Formative** evaluation produces information which is fed back during program operation for the purpose of ongoing program improvement. **Summative** evaluation, on the other hand, may withhold feedback to maintain stable program operations and emphasizes relatively final judgment of program effectiveness, usually rendered after the fact. Summative evaluation is most relevant to evaluation research; formative evaluation to what we describe as program evaluation. Formative approaches seek to strengthen ongoing programs, hoping thereby to enhance the ultimate effects. **Formative approaches are most relevant to the majority of local prevention programs.**

Levels of Evaluation: Process, Outcome, and Impact[5]

The Process level: Here the inputs and activities of the program itself are considered. This is useful for formative program evaluation, describing staff and participant characteristics, interactions, and transactions; for understanding the program's theory; for needs assessment; and for detailed description and documentation of the services actually rendered. Process data is useful both in day-to-day management and planning activity, and in disseminating a program model for possible implementation or replication at another site.

Process evaluation often focuses on **effort.** The quantity and quality of activity which takes place is assessed, regardless of output or effectiveness. Focus is on programmatic inputs, such as the number and types of clients and staff, resources expended, extensiveness of services provided, and similar indicators of what a program did. Evaluation approaches which attempt to measure effort are known as "program monitoring." Such approaches measure the program's capacity for success and documents level of activity. **All prevention programs should establish program monitoring systems.**

Outcome evaluation examines the attainment of program objectives related to short- and long-term change in participants' behavior, attitudes, knowledge, or level of problems or dysfunctions. Attaining satisfactory outcomes is the primary or ultimate reason that programs are established, and thus deserves attention in most efforts at program evaluation.

Smoking prevention programs would have successful outcomes if a meaningful percentage of participants did not begin to smoke in the future. A genetics counseling program could be judged successful if it resulted in a lower than predicted rate of genetic abnormality among offspring of its participants.

Adequacy of outcome, the degree to which the obtained outcome is adequate to the total need, can also be considered. While some programs may be highly effective for those few persons receiving their services, the cost may be so high as to prohibit wide scale implementation, in spite of a great deal of need. In such cases, adequacy of outcome would be very low, even though outcome was very positive.

Efficiency may also be looked at as part of an outcome study. This involves the consideration of alternative strategies for minimizing costs (effort) while maximizing outcome and adequacy. Could the same level of outcome be accomplished at less cost in terms of money, time, personnel, and recipient time, energy, and inconvenience? Cost-benefit and cost-effectiveness research measure program efficiency.

Impact evaluation refers to the assessment of longer term, generalized results of program operation. This relates to the community-wide needs and problems which led to the establishment of the program. Impact evaluation is thus especially relevant to prevention efforts which often have community-wide influence as a goal. Impact evaluation helps to ascertain whether the total effect of the program (especially its "spinoff" results) has been beneficial. In the area of alcohol abuse prevention, impact evaluation may seek to determine whether the program has reduced the incidence and prevalence of alcohol abuse among youth in the community, whether more alternatives are available and used, whether there are fewer drunk driving incidents involving youths, or other issues related to the program's goals and objectives.

The key difference between **outcome** and **impact** evaluation is that outcome considers only the direct program participants, while impact assesses total effects on a community, neighborhood, or other larger target population. Due to its complexity, impact evaluation is covered only in passing in this guide.

All three levels do not have to be implemented in any given evaluation effort. In fact, trying to establish all three levels is unrealistic, inappropriate, and counter-productive in most program evaluation efforts.

Consideration of these distinctions hopefully has lead you to an expanded sense of the range of possibilities in the evaluation enterprise. Evaluation is not a single method, design or approach, but a smorgasbord from which to pick and choose as appropriate to accountability requirements, your information needs, and the available resources. While a comprehensive evaluation will provide answers to all parts of the question *Who does what to whom with what results at what costs?*, the rigorous evaluation called for to completely answer this question is beyond the resources of most local prevention programs. Rather, you must decide which parts of this question are most relevant and feasible to answer in any given evaluation effort.

CHAPTER 2
WHY EVALUATE PREVENTION PROGRAMS?

As programs and services become institutionalized, the demand for formal evaluation diminishes. Rather than using meaningful evaluation as a basis for support, institutionalized organizations develop a sense of confidence, good faith, and political support which minimizes inspection and evaluation both internally and externally.[6] Funders and constituents usually take for granted the effectiveness of such organizations and programs, and rarely require rigorous evidence of success.

However, new programs, services, or organizations which have not established themselves over time and which are not institutionalized are often required to provide rigorous evidence of their effectiveness. Prevention programs especially face this dilemma as they attempt to legitimate themselves. Prevention people often appear to talk a different language than do persons in established remedial service areas, and the programs rarely incorporate the expected common elements of most human service programs. Furthermore, the **constituency** for prevention programs is relatively politically weak or unorganized, particularly in comparison to easily identifiable disability groups such as recovering alcoholics, parents of developmentally disabled children, or day care lobbying groups. **Funding** for prevention has not, until very recently, been routinized or taken for granted as it has been for intervention and treatment programs. In this context, program evaluation provides an opportunity to demonstrate program value, develop constituencies, and justify funding of successful strategies. A further reason for evaluation, of course, is that it is required by most prevention grants and contracts.

It is also important to evaluate prevention programs to **determine which program models are, in fact, effective.** Even programs which "seem to work," are liked by participants, and appear to be a good thing to do can have negative effects. For example, a summer jobs for youth program, meant to prevent delinquency, had the effect of increasing delinquent behavior among participants during the school year following program involvement.[7] The Cambridge-Somerville Youth Study, designed to prevent delinquency, delivered a wide and extensive variety of preventive services. A thirty-year follow-up of participants found significantly higher rates of "undesirable outcomes" (early death, alcoholism, mental disorder) among former participants than among controls.[8] Without follow-up and outcome evaluation, both of these projects would have appeared to be effective and to have had high levels of participant satisfaction.

With limited funding available to the human service system as a whole, and with widespread distrust and skepticism concerning the value and effectiveness of prevention programs, evaluation is particularly called for. Those seeking funds to provide services to the persons already suffering from the problems prevention tries to prevent have a legitimate point when they fear a loss of funds needed to deal with obvious and immediate problems. The prevention field must, in this context, demonstrate that funds used in prevention are well spent in terms of their impact on future problems. Conversely, treatment programs should also be required to demonstrate the long-term benefits and cost-effectiveness of their services. Thus, evaluation is needed to legitimize prevention activity and demonstrate its efficiency and cost-effectiveness, and to eliminate ineffective approaches.

Uses of Evaluation

The most common use of evaluation at all program levels is that of **accountability.** Most grants currently require some form of evaluation as part of an accountability system; all funding sources seek some level of information on at least the process level. These external demands range from simple "head counts" (how many clients were served) to extensive demands for evidence of effectiveness. Internally, program managers usually also seek some evidence of performance ranging from simple demands for documentation of effort expended (e.g., staff time spent) to complex systems of performance effectiveness monitoring.[9] Monitoring systems can serve to provide these accountability data.

Another purpose program evaluation can serve is that of **program planning and development.** Needs assessment, often considered a part of program evaluation, determines the extent of a community's need for possible programs.[10] Data generated from needs assessment studies can be applied on many levels, including use by existing programs, funders, and planning agencies in determining future directions and identifying gaps in existing services. Program evaluation data from ongoing monitoring systems can also assist in program development and planning by indicating trends. A simple system which collects data on participants can point to changes in participant characteristics over time, and programs can be fine-tuned or altered to better meet the needs of those clients. For example, data regularly collected and analyzed from an alcohol and drug abuse prevention agency providing an information and referral service showed an increase in the number of inquiries regarding a group of drugs with abuse potential. The agency was able to respond by increasing its base of information regarding this drug, as well as incorporating educational materials regarding the substance into its school-based prevention efforts.

The evaluation process also **forces program personnel to systematically examine what they are doing and why.** All programs are based on some sort of conceptual model or theory; however, this model is usually implicit. As the evaluation process gets under way, personnel are forced to make this explicit, leading to a better understanding of their own program, realization of possible contradiction between various parts of the program, and further development of the program as a consequence.

Evaluation can, in other ways, also **assist in program improvement.** With regular feedback about program operations and results, administrators and practitioners can take rational action based on systematic knowledge rather than impressions. The data can be of many types, including simply identifying problem areas, pointing out overlooked needs of participants, documenting inadequate staff performance, discovering environmental difficulties that impact on the program, and pointing out negative or unanticipated outcomes. Program strengths that can be capitalized on may also be identified.

Without systematic evaluative data, decision-makers and practitioners must function on the basis of often mistaken impressions of what is going on. A relevant knowledge base, gained from evaluation data, can serve to reduce uncertainty, point to possible improvements, and at times even provide predictions concerning likely results of alternative program improvement strategies.

Program evaluation, particularly on the process level, can also be used in the **dissemination of program models** for replication. The program, as implemented, can be described, problems encountered and solutions attempted pointed out, and recommendations for implementation at other sites made. Thus, the experiences of one program (including guidelines, suggestions to be followed, and some expected results) become available for others to use.

Another use of program evaluation is to **test theories of prevention** implemented by a particular program approach. While this purpose is best served by an "evaluation research" model, it can be a part of a program evaluation attempt. The assumptions upon which a program is based may also be tested.

Summative uses are best accomplished through evaluation research models, but may also be addressed in program evaluation. These uses include making decisions concerning whether to continue or discontinue a particular program, whether to add or drop specific strategies or techniques, whether to institute a model program elsewhere, or how to allocate resources among competing programs.[11]

Given the political nature of much program evaluation, misuse of evaluation also occurs regularly. Calling for an evaluation can be used as a means to delay or postpone a decision, to duck responsibility, to justify an unpopular decision already made, or to co-opt opponents by involving them in the evaluative process. Evaluation tends to lend legitimacy to an effort (regardless of the actual findings), and thus may be used for purposes of enhancing the credibility of a questionable approach, and for general public relations.[12] Evaluation may be used to support the existing way of approaching a problem, or it may be used to recommend changing only marginal issues rather than basic approaches.[13]

Evaluation data tend to be used most often on a **conceptual level;** to clarify issues, redefine problems, and to redirect the attention of users to other issues. This can be described as "evaluation as education,"[14] which results in more sophistication about the appropriate questions to ask and which stimulates critical self-awareness among practitioners. For example, practitioners may initially be concerned with the overall effectiveness of their program. As evaluations show mixed results, the question may be redefined as "effective for which sub-groups of program participants?"

A great diversity of purposes can be served by evaluation. The potential use is related to the concern of those requesting an evaluation, be they local programs, funders, state or national policy makers, or others. No one evaluation or ongoing evaluation system can serve all potential uses, and **it is thus crucial at the beginning to determine what are the intended or likely uses, and who the primary users are.** Based on this, appropriate evaluation questions and issues can be identified and focused, with participation of all involved, leading to a greater likelihood that usable findings and data will be generated.

CHAPTER 3

GETTING STARTED: PLANNING PROGRAM EVALUATION

The previous chapters provided an overview of some purposes and options available for the evaluation of prevention programs. We have stressed that doing evaluation is not a simple matter of applying a set of predetermined techniques—rather, **evaluation must be developed and tailored to the unique circumstances of the program under scrutiny.** Of paramount concern throughout this planning process is the usefulness of the evaluation to the key decision-makers and others who are involved.

Exhibit 3.1 provides an overview of the recommended steps in planning and carrying out any program evaluation. The evaluation model presented here emphasizes maximizing its usefulness to those who will use the results.[16] Since the audience of this guide is primarily local programs, we concentrate on evaluation as a tool for improving local prevention programs, providing accountability information, and establishing data on prevention strategies which are relatively successful. Regardless of the focus of the evaluation, these steps should be followed.

Step 1: Identify and Organize Relevant Information-Users and Decision-Makers

Your development of an evaluation strategy should always involve the evaluator *and* key staff. (The "evaluator," in small programs with limited resources, may well be a designated staff member or administrator. The term is used throughout this booklet to signify the key person responsible for overall conduct of the evaluation.) Key staff should include the key decision-makers who will use the evaluation, as well as some staff who deliver or supervise services.

This group of decision-makers and information-users (henceforth "key users group") should be made up of people who (1) want and can use information, (2) who have questions they want answered, (3) who will act on the basis of evaluative information, and (4) who will devote time to and share responsibility for the evaluation. The group should meet regularly with the evaluator and be of a workable size (3 to 7 persons is a reasonable range). This key users group, rather than the evaluator or an external bureaucracy, is seen as owning the evaluation.

Exhibit 3.1: ***Steps in Planning and Carrying Out an Evaluation***[15]

1. Identify and organize relevant information users and decision makers.
2. Identify and refine the relevant evaluation questions.
3. Specify program goals and objectives.
4. Select evaluation methods appropriate to the evaluation questions.
5. Develop an evaluation plan.
6. Pilot test the evaluation system.
7. Implement the evaluation plan.
8. Summarize, analyze, and interpret the data, ideally with the involvement of information-users and decision-makers.
9. Disseminate and use the findings from the evaluation.

Step 2: Identify and Refine the Relevant Evaluation Questions

After establishing the key users group, begin the actual planning process. The members of the key users group, with the evaluator, must first determine the broad purposes of the evaluation and choose from among the many alternatives available. Is the evaluation to be formative or summative? Should it emphasize process issues, outcome issues, or impact issues? Which parts of the prevention program should be examined? Is simple monitoring of participants and services all that is feasible?

To guide the process of selecting and developing appropriate evaluation questions, the evaluator ought to pose a series of diagnostic questions such as those in Exhibit 3.2. Explicitly focusing on these can assist the program staff to better conceptualize their services, the constraints they operate under, and the theory behind the program. This may lead to program development and improvement without the actual collection of any evaluative data.

When a new evaluation is implemented in an existing program, your ongoing data needs and systems must be considered. To the extent possible, an evaluation should use existing data systems and generate required data and reports as a by-product. Reduce duplication in data collection and reporting as much as possible. Take external reporting and evaluation requirements into consideration, even though the basic concern is usefulness to the key users group.

After exploring the issues raised in Exhibit 3.2, ask the key users to complete the following statement:

We need to know:

Because we need to (decide/demonstrate/account for):[17]

You can enhance the usefulness of your findings by structuring a group work session in which a series of these statements is developed, and then selecting the most important or relevant questions as the focus for evaluation.

Your final selection of evaluative questions should be based on several criteria, including the possibility of bringing data to bear on the question, the existence of more than one possible answer to the question, and the extent to which answers to the question(s) will be useful to decision-makers.

Step 3: Specify Program Goals and Objectives

Among the most often asked evaluation questions is "Is it effective?" Is the program attaining the desired outcomes among its participants or other target group? In order to answer this question, you must clarify what the program is attempting to accomplish—its goals and objectives. Programs which do not have well

Exhibit 3.2: *Diagnostic Questions in Identifying Relevant Evaluation Questions*

- What is the nature of the program? Specifically, what is its:
  - —Assessed need
  - —Implied or stated conceptual or theoretical model
  - —Perceived mission
  - —Anticipated benefits, goals, and objectives
  - —Intervention/service delivery approach
  - —Target population(s)
  - —Extent of exposure to target population
  - —Resource base and stability
- Who needs to know what, and why?
- What data are absolutely required for internal and external reporting?
- What data are currently available or collected, and in what form?
- What are the major issues being faced by the program?
- What is the community context in terms of values, culture, and politics?
- What external constraints (standards, regulations, etc.) must the program demonstrate compliance with?
- What resources (time, money, expertise) are available for the evaluation?
- Has the program model been adequately tested elsewhere?

thought-out goals and objectives are likely to meander. Without a central focus, they are not amenable to evaluation.

Cantor *et. al.* suggest the following steps to develop outcome objectives:[18]

Step One: Write Down General Goals.
Goals are broad statements of program purpose, such as "the prevention of child abuse," "improvement of family life," or "to improve future occupational choices of participants." Given that goals are broad statements of purpose, most programs will have only one or two goals.

Step Two: List Possible Indicators of Goal Achievement.
Indicators are specific and observable changes in attitudes, knowledge, or behavior linked to goal achievement. At this step, a large and exhaustive number of indicators of goal achievement should be generated. Examples of indicators include "fewer reports of possible child abuse to county social services from the target families," "a reduction in the number of family arguments reported," "a reduction in the rate of school dropout among participants," or "an increase in ability to turn down a cigarette."

Step Three: Select the Best Indicators.
Choose the best indicator(s) from the list you generated, based on their significance and relevance to your target population, their importance to decision-makers, the ease with which they can be measured, and your program's ability to have an impact on them.

Step Four: Translate the Indicators into Measurable Objectives.
Measurable objectives should contain a statement of the indicator and target population, a time frame, and either a statement of the proportion of the target population expected to show change or the amount of change expected on the indicator. Alternative formats are:

A. _____ (By When?), _____ (Percent or % Change) of _____ (Who?) will _____ (Indicator).

B. _____ (By When?), _____ (Who?) will _____ (Percent or % Change) of _____ (Indicator).

Examples include: "By 12/31/84, 75% of the families will report a decrease in the number of serious arguments in the family." "By the end of the school year, less than 5% of the juniors and seniors at high risk of dropping out will do so." "By the end of the program, the seventh graders will show a 10% increase in the number of socially acceptable ways to refuse a cigarette which they identify."

Assigning percentages to these changes is difficult. The assigned percentage is considered a minimum performance standard, based on the key users group's perception that if we accomplish this amount, the program has succeeded. In setting objectives, you should consider past experience, research and evaluation literature, the strength of the program, and the difficulty of changing the behavior of concern.

With a clear statement of goals and outcome objectives, the stage is set for evaluation of outcomes. Program personnel are also better able to communicate what their program is attempting to accomplish to outside audiences.

Step 4: Select Evaluation Methods Appropriate to the Evaluation Questions

Having determined the focal evaluation question and the program's goals and objectives, select a method of obtaining the necessary data to answer the questions posed. As indicated earlier, a wide variety of evaluation techniques and approaches is available. There is no one evaluation method. The wide range of options available may make this choice a difficult issue; this is a technical area requiring more expertise than a brief guide can provide. In many cases **you may require consultation from a specialist in evaluation and research methodology at this point.**

The range of options available regarding general **methodology** are outlined in Exhibit 3.3. While some approaches specifically require the collection of new data, others can use existing information (e.g., a retrospective study) or simply review the program as described in existing documentation and by key staff members (e.g., the review by experts who evaluate and suggest improvements based on standards in the field and their experiences with similar programs).

Your choice will depend upon the evaluation questions asked, the goals and objectives established, the availability of existing data, and the resources available. If the evaluation question is a monitoring one (e.g., How many people with what characteristics are participating in the program?), it is desirable to use a pre-structured approach, focusing on future participants in the total program and collecting data from all participants. On the other hand, if you ask a question such as "How effective is the program in promoting healthy lifestyles?" the choice of general methodology is much more complex.

Once you determine a general methodology, you must consider options in data collection (Exhibit 3.4). Again, the choice may be simple or difficult, depending upon your concerns. A process evaluation (see Chapter 4) to develop a detailed operational description of the program might use narrative (qualitative)

Exhibit 3.3: *Options Regarding General Evaluation Methods*

- Pre-structured or allowed to vary with accumulating data? (Fixed or Dynamic?)
- Concern with past participants or future participants? (Retrospective or Prospective?)
- Focus on total program or on specific parts of the program? (Holistic or Component Focus?)
- Test specific hypotheses about the program, or develop generalizations based on the data? (Deductive or Inductive?)
- Compare groups receiving the program to similar groups who do not get the program, or look only at program participants?
- Use data collected regarding program services and participants, or a general review by experts?

data, collected through staff and participant interviews, observation, and open-ended questions. There would be little immediate concern for developing specific indicators or scales. An outcome study, on the other hand, would probably rely on numerical (quantitative) data, collected by interview or questionnaire from a random sample of participants, using structured questions. Instruments and indicators or scales depend upon the outcome of concern. While a smoking prevention program may consider a single indicator (e.g., any smoking within a one-year period), a family life enhancement program might use a battery of multiple item scales (some of which are pre-existing) attempting to measure levels of family communication, family cohesiveness, and satisfaction with family life. To deal with complex issues, such as "family life enhancement," you will probably need external consultation.

Since instruments (i.e. data collection forms) are often needed by many programs, a checklist is provided in Appendix B to use in developing your instrument. It is preferable to use **existing** measures of known validity and reliability (technical quality). However, take care to assure the relevance of the measure to your population.

Issues to Consider in Selecting Existing Instruments:

1. Do the items reflect the key user's concepts of what they want to measure?
2. Are responses to the items likely to change as a result of the program?
3. Is the length of the instrument and the time it takes to complete it appropriate to the time available for respondents?
4. Is the scale **reliable** and **valid** when used with populations similar to yours? (**Reliability** refers to the internal consistency of items making up the scale and/or its ability to yield the same results when repeatedly applied in the absence of changed behavior or attitudes. **Validity** refers to whether or not the instrument measures what it claims to measure.)

Exhibit 3.4: *Data Collection Options*

- Numerical or narrative? (Quantitative or Qualitative?)
- From existing records, by observation, by interview, or by questionnaire?
- From samples or entire groups (populations)?
- Open-ended or structured questions?
- Existing instruments or ones developed for your own purposes?
- Single-item indicators or multiple-item scales?

Since data is unlikely to be used unless it is believed by the users, the evaluator and key users group should review all aspects of the evaluation method to determine its limitation and believability. Consequently, one consideration in selecting appropriate instruments should rightfully be "face validity"—the instrument looks like it measures what it is claimed to measure.

Multiple Methods

We also encourage use of multiple methods of obtaining similar or complementary information. If results converge, this further adds to the believability of your information.

For example, in an evaluation of a cross-age peer education program to prevent alcohol and other drug abuse, many of the high school age peer facilitators reported on a post-test that the program had led them to more responsible drug use. The process evaluation of this program included interviews of many of these students and independently found movement in the direction of more responsible use patterns, adding confidence to the conclusion.[19]

Relevance of Method

Before making a final selection, review the proposed method to determine whether it will, in fact, produce the data necessary to answer your key user group's questions. Evaluations with complex, rigorous meth-

odology can easily be irrelevant to the evaluation questions being asked. Prevention programs must be cautious about implementing evaluation designs merely because they represent the "state of the art." The true issue is whether the method is relevant and adequate to answer the questions being posed.

Feasibility

Finally, examine the **feasibility** of proposed evaluation methods. Are there sufficient **resources** available to conduct the proposed evaluation? Consider all resources including:

- Budgetary allocation
- Staff time, both to conduct the evaluation and to participate by providing data (filling out forms, participating in interviews, etc.)
- Necessary expertise (especially for technical aspects of evaluation, such as instrument development, sampling, statistical analysis, computer work)
- Availability of data entry and computer services, and budget line item to cover this
- Clerical time for typing forms, mailings, data coding, report preparation
- Duplicating costs for interview forms and questionnaires

Beyond the availability of resources, is there sufficient time to carry out the evaluation as proposed? A one-year outcome study, for example, cannot be done under a one-year grant, since it will usually take a full year just to collect baseline data from a sufficient number of participants who are to be followed the next year, not to mention start-up time. A decision which must be made in three months cannot rely on evaluation data which will take six months to collect.

Another feasibility issue to consider at this step is the burden which may be placed on persons asked to provide data. We tend to request more information than is absolutely necessary in most evaluations; when this puts an undue burden on respondents, the response rate or accuracy will be reduced.

Finally, the evaluation method must be appropriate to the developmental stages of the prevention program. Many programs require start-up times of six months to two years or more before they become established, work out flaws, and settle down to a routine and relatively stable model. To insist on outcome or impact evaluation from the start may doom your program to **apparent** failure. This is particularly the case with community-based prevention programs, which may need to spend a year or two in community organizing and community development, before tangible preventative results can be hoped for. Thus, the proposed evaluation method should also be **feasible** in terms of the program development time table and process.

Step 5: Develop an Evaluation Plan

Having identified information users, determined the evaluation questions, and tentatively selected evaluation methods, next develop a complete evaluation plan. As with each of the prior steps, the key user's group should be involved, at least in the final negotiation and revision of the plan. What emerges from this stage will serve as a blueprint for the evaluation.

The plan's formality will depend upon the scope of the evaluation; for small scale efforts, the plan may only be a rough set of notes developed by the evaluator. At other times, a complete written document should be developed and distributed to all those involved.

The evaluation plan should include:

- A statement of the evaluation question(s).
- A detailed description of the evaluation method: what data will be collected, from whom, using what techniques?
- Who will be responsible for each evaluation task?
- A statement of the conditions under which data collection will take place, and methods to be used to standardize data collection.
- Plans for regulating the flow of evaluation data.
- Plans for summarization and analysis.
- Plans for protection of subjects, including obtaining informed consent and limiting access to identifying information.
- Time lines for all phases of the evaluation, including data collection, analysis, reporting, and pre-set checkpoints with key user's group.
- Plans for use of the information, which explicitly take into account the level of certainty which can be placed on the findings, given the trade-offs which have been made and the problems of design and instrumentation.

Once drafted, negotiate the plan with the key user's group before it is completed. Again, the evaluation plan should be seen as this group's property, building their commitment to use of the evaluation by giving them ownership.

Evaluation Roles

The statement concerning who will be responsible for the various tasks should include all aspects of the evaluation, including drafting the instruments, collecting and analyzing data, and preparing reports. Often in small programs, practitioners are asked to serve as evaluators as well. The role of the evaluator has been described as that of a "sympathetic skeptic,"[20] who must raise tough and critical questions about program processes and outcomes, and objectively bring data to bear. However, the practitioner must believe in what s/he is doing in order to function effectively. It is diffi-

cult for practitioners to examine their own work with the skepticism necessary to do a good job of evaluation, and yet continue to simultaneously function as effective prevention workers. Thus, assigning evaluation duties to practitioners should not be done without careful consideration of the options and possible effect on the program as well as on the evaluation.

The selection of an outside person to conduct an evaluation is also difficult; one would want a person sympathetic to the program rather than one who will simply do a "hatchet job" as a result of the evaluator's inherent skepticism.

Informed Consent and Protection of Confidentiality—Ethical Issues

Persons providing data for evaluation purposes have a right to know what the data will be used for, how their participation will potentially benefit and/or harm them personally, and a right to refuse without adverse consequences. (This is known as "informed consent"; specific regulations apply to situations in which federal money is used to fund data collection.) To assure confidentiality, access to data should be carefully protected. For example, in school-based prevention programs which involve pre- and post-testing, protecting confidentiality may be a problem. If teachers are involved in collecting forms after the students have filled them out, confidentiality may be broken because the teacher recognizes a student's handwriting and reads his/her responses. Take care to establish and maintain confidentiality, and to avoid situations which may have the potential of violating confidentiality.

As part of the time lines established for the evaluation, you should include pre-set check points with the key user's group to maintain their interest and involvement, deal with problems which develop, and negotiate changes which need to be made in the plan over time.

Step 6: Pilot Test the Evaluation System

Once an evaluation plan has been developed, but before full implementation, pilot test the entire evaluation system. Try out all instruments (forms, interview schedules, and questionnaires) on several cases. This will allow for revision and elimination of problems before all the data are collected. Also test your plans for recording or coding the data in order to avoid major problems later.

During the pilot phase, consider the following issues:

- Did the respondents understand the questions? Were they willing to respond?
- Did the interviewer or respondent have trouble reading the questions as written?
- Were there inconsistencies or ambiguities in interpreting (coding) any of the data?
- Did data collection take too long?
- Was the format confusing?
- If a new form has been introduced as a standard procedure, did all who filled it out understand the items in the same way? Was anything missing which will interfere with another person's work? How will it effect other systems (fiscal, administrative)?

Inevitably, some alteration will be needed following a pilot. Although some see this step as tedious, these alterations will prevent many larger problems which would occur later if no pilot were conducted.

Step 7: Implement the Evaluation Plan

At this point, introduce the evaluation plan as an ongoing routine. It is important that the person responsible for the evaluation monitor all aspects. There are many areas where procedures can break down, with serious potential for jeopardizing the evaluation and possibly harming the program's recipients. This monitoring can, for the most part, consist of informal spot checks or procedures.

Potential problems during implementation include:

- Breakdown of participant assignment to experimental and control groups.
- Failure of field workers to collect and record the data they are expected to.
- Reconstruction of data long after-the-fact, with inherent inaccuracies. This is often a major problem with practitioner's paperwork.
- Breakdown of informed consent and confidentiality protection procedures.
- Inaccurate or sloppy coding of data.
- Biased sampling, particularly of program recipients.

You can avoid some of these problems by requiring that all data forms be turned in (and briefly examined) within a short time after completion, with sanctions built in to enforce this. Others can only be detected and solved with close involvement of the evaluator, who should be familiar with the procedures actually used in all aspects of the evaluation.

Step 8: Summarize, Analyze, and Interpret the Data

Depending upon the approach used, data summarization, analysis, and interpretation will either be an ongoing process which occurs simultaneously with data collection or will be a separate step which occurs after all data have been collected.

Your evaluation plan, if comprehensive, will have specified the appropriate analysis. It is beyond the scope of this guide to deal in-depth with analytic techniques.[21] If sophisticated quantitative analysis is required, most prevention programs will need specialized assistance.

Quantitative Summarization and Analysis

Most summarization can be done by hand tally and pocket calculator if the number of cases and variables is small. This is best done by transferring all data onto a tally sheet, so that columns can be easily summarized. Simply frequency distribution (giving the number and percent of cases falling into each category) are most common. Cross tabulations of items may also be produced, which can provide for basic comparisons between two groups of recipients, or show the relationship between two variables. Finally, measures of central tendency (mean, median) can be calculated and compared or used descriptively. (These simple techniques are often all that is needed; more complicated bivariate and multivariate approaches such as analysis of variance, contingency table analysis, multiple regression, and analysis of covariance are beyond the scope of this guide.[22])

A simplified example to demonstrate this process is from hypothetical case records of a drop-out prevention program targeting high-risk high school students. The program kept a standard record (monitoring form) on each participant, which included an ID number, the participant's sex, his or her age at entry into the program, and his or her outcome (graduated, passed the G.E.D., or dropped out). For analysis purposes, these items were extracted and arranged on tally sheets (Exhibit 3.5), one case per line, so that it was easy to summarize by looking at columns of data rather than shuffling through many forms. Extracting data from the tally sheets produced the frequency distribution of outcomes (Exhibit 3.6), indicating an overall dropout rate of 14.3%. Cross tabulation of participant sex by outcome yielded the cross tabulation in Exhibit 3.6, which indicates that female students were slightly more successful than males—17% of the males dropped out, but only 8% of the females. (Statistical significance of this result could be determined by using a Chi-square test of significance.)

If the number of cases is large, obtain assistance in computer analysis. This is most readily available from local colleges and universities, as well as from some school boards, insurance companies, private consultants, or municipal governments. **Input from a computer specialist must be attained at the pilot phase or earlier,** to assure that the data are collected in a proper format to allow for analysis.

Exhibit 3.5: *Example of Data Tally Sheet for a Dropout Prevention Program*

	ID Number				Sex Code		Age			Outcome Code	
Col. #	1	2	3	4	5	6	7	8	9	10	11
	0	0	1		1		1	6		1	
	0	0	2		2		1	7		2	
	0	0	3		2		1	7		1	
	0	0	4		2		1	6		3	
	0	0	5		1		1	7		.	
	.	.	.		.		.	.		.	
	.	.	.		.		.	.		.	
	.	.	.		.		.	.		.	
	0	7	0		2		1	7		3	

Codes:

Sex:
1 = Female
2 = Male

Outcome:
1 = Graduated
2 = GED Passed
3 = Dropped Out

To obtain *frequency distribution* of sex:
Count the number of 1's and 2's in column 5.

To obtain *cross-tabulation* of sex by outcome:
Count separately the number of 1's, 2's, and 3's in column 10 for each 1 in column 5; repeat for each 2 in column 5.

To obtain *mean (average) age:*
Add the numbers (ages) in columns 7-8; divide total by the number of cases (n = 70).

Exhibit 3.6: *Examples of a Frequency Distribution and Cross Tabulation for a Dropout Prevention Program*

"Overall Participant Outcomes in the High School Dropout Prevention Program During 1982-83"

Outcome	Frequency	Percent
Earned H.S. Diploma	55	78.6%
Passed G.E.D.	5	7.1%
Dropped Out, No G.E.D.	10	14.3%
Total Participants	70	100.0%

"Participant Outcomes by Sex, 1982 – 83"

	Males	Females	Total
Earned Diploma	35(76.1%)	20(83.3%)	55(78.6%)
Passed G.E.D.	3(6.5%)	2(8.3%)	5(7.1%)
Dropped Out, No G.E.D.	8(17.4%)	2(8.3%)	10(14.3%)
Total	46(100.0%)	24(99.9%)	70(100.0%)

Qualitative Analysis

Qualitative data renders a narrative analysis which may take many forms. One approach is to reproduce the data without comment, for example, by participants' responses to open-ended questions on satisfaction surveys. The data may be organized by topic, for example, by pulling together all responses from interviews which deal with perceived effectiveness of the program. (It helps to have an extra set of interview transcripts, field notes, and other recorded data to use for cutting and pasting.) Data may also be organized into descriptive case studies, for example, describing in-depth the experiences of several typical program recipients. Typologies (which isolate and categorize the essential characteristics of common types) of events, programs, recipients, or their experiences may also emerge from qualitative analysis, which can be refined as relevant to the evaluation question.[23]

Analysis as an Interactive Process Involving the Users Group

Interpretation of the data should be a joint effort involving both the evaluator and the key users, in order to enhance usefulness, avoid unexpected surprises in a final report, and obtain added perspectives on the meaning of the data.

After the users group initially considers the data, further analysis is likely to be required. Data analysis is a process in which the answer to one question often leads to the posing of another.

For example, in evaluating an alcohol and other drug abuse prevention program using a pre-test and post-test without a control group, it was found that seventh graders reported increased use of marijuana and increased intentions to use in the future after exposure to the program. This raised the further question of (1) whether the students who increased on this variable differed from those who did not in other ways, and (2) whether students *not* exposed to the program also increased on these variables. (If an experimental design had been used in the first place, this would not have become an issue.)

Interpretation is rarely straightforward and obvious. Most studies have both negative and positive findings; they also usually have methodological limitations or problems which call into question the certainty of results. Thus, your interpretation should be specific, focusing on concrete aspects of the program, rather than characterizing the whole program as "successful" or not. Report strengths and weaknesses of the evaluation methods fully, so that accurate judgements can be made of the degree of confidence to be placed on the results.

While most evaluations conclude with a final written report, this is not always necessary. Verbal reports, with supporting table, graphics, and charts, or case studies and organized qualitative responses may be sufficient for the needs of the program. Most, however, will require a final report summarizing the details of methodology, findings, and interpretations. Several versions of the report may be desirable, perhaps including a brief, simplified version for widespread dissemination.

Step 9: Disseminate and Use the Findings

Your key users group should proceed next with translating the results of the evaluation into program operations and policy. This direct application does not often occur as a result of most evaluations, however. Use of evaluation and research findings more commonly is a gradual, cumulative process in which the questions are changed, conceptualizations are altered, and which only has an indirect impact on policy decisions. We must realize that evaluative data is rightfully only one input among many into program decision-making, and we should not expect more direct impact than is warranted.[24]

One major area of use is in **replication.** Successful prevention programs are in demand; findings from evaluation can assist in legitimizing specific models and in making recommendations for their most effective implementation elsewhere.

CHAPTER 4

PROGRAM MONITORING AND PROCESS EVALUATION: A COMPREHENSIVE VIEW

Process evaluation is most often viewed and practiced as a mere exercise in head counting—of program recipients, of meetings held, and of staff activities. For accountability purposes, this may be necessary to document that program funds were spent as intended and that services were delivered. However, this view often limits the utility of the evaluation.

A more effective use of process evaluation is a "comprehensive" one that **stresses an analytic examination of program operations in their totality.** Process evaluation seeks to answer the questions: "What is this program, really?"; "What services were provided to whom?"; and "Why did we obtain the outcomes we did?" Any systematic collection and analysis of data about program operations is a form of process evaluation, although most process evaluation will not reach the ideal comprehensive potential.

Process evaluation often is the most immediately useful and feasible approach in an evaluation effort. It provides more information of immediate relevance than most outcome or impact studies. For programs with limited evaluative resources, establishment of a monitoring system combined with some other form of process evaluation is recommended as the place to start. As with any evaluation, the specific focus will depend upon the questions being asked.

While process evaluation deals with the actual delivery of prevention program services, it does not indicate success or effectiveness; only the capacity for success. However, demonstration of successful outcomes may be beyond the resources of your program. Outcomes may be so far in the future that measurement is impractical, or successful outcomes may already have been demonstrated convincingly for the same program model. Some prevention programs take several years to implement and develop fully, during which time little or no measurable outcome can be expected. Others are oriented to a process—such as community development—which has very broad, generic, difficult-to-measure outcome goals. In all of these cases, process evaluation may be the only feasible evaluation approach.

Process evaluation may incorporate a variety of components. **All programs should have some type of ongoing monitoring system.** Beyond this, you can pick and choose from these options, depending upon the evaluation questions being asked.

1. Ongoing Monitoring Systems are geared primarily to documenting various aspects of your process or effort. This most often takes the form of a service or case record, with data on case characteristics and flow (demographics, referral source, prevention activities involved in, disposition) as appropriate to the program. Most programs maintain some form of case record, but often these records are not standardized, are inaccurate, and difficult to summarize. By systematizing and standardizing these records, and developing a data retrieval system, a great deal of valuable evaluative data can be generated. A monitoring system enables a program to accurately answer the questions of: "How many people did you serve last year?"; "What program services did you provide?"; and "What are the characteristics (age, sex, income level, etc.) of your participants?" By aggregating client data over a specified time frame (preferably as an ongoing system), clinicians and decision-makers can answer these and other relevant questions about their programs. Demographic and assessment data can indicate whether your intended target population is in fact being reached, and, by inference, whether the services available are appropriate. Trends in changing characteristics can be assessed and planned for. Referral source data can be used in examining the program's relationship to the community and can suggest necessary outreach efforts. Data on services provided may indicate

that many participants have more or less than the planned exposure to the program, leading to reexamination of the nature of services provided.

Monitoring of Project NEWS, a community development program for health promotion and substance abuse prevention, is illustrative. The project had originally intended to reach mid-teens and families who had not previously been involved in neighborhood center activities. Monitoring revealed that the actual participants were largely pre-teens and already involved families. This suggested making the program content more relevant to pre-teens, as well as increasing efforts to reach uninvolved potential participants.[25]

An ongoing monitoring system is also indispensable in the accountability process. At a minimum, most funding sources want to know the number of recipients, basic demographic characteristics of recipients, and the number of contacts provided. A basic monitoring system can be structured to provide these data with little effort and expense.

In programs with identifiable participants who are involved over an appreciable length of time (as in ongoing education programs and Family Enhancement Program group sessions), the monitoring system should be developed on a participant record basis. Records can be kept on notecards or on larger forms, one card or form per participant. Determination of what data to include should be based on the nature of the program and its information needs; Exhibit 4.1 lists some of the most common options. Developing the participant record format should follow the steps outlined in Chapter 3 and Appendix B, which are relevant to any data collection instrument.

If participants are not readily identifiable, or are involved for brief or sporadic time periods, monitoring data should be collected on a service delivery basis. Prevention programs such as community organization and education programs, mass media campaigns, health fairs, or Positive Youth Development lend themselves best to this approach, since the participants are often unidentifiable. Monitoring of service delivery uses the activity as the unit of analysis; a brief form is completed for each distinct service activity conducted.

The data on such a monitoring record may include:

- Site of service activity
- Nature of the activity or service provided
- Amount of time spent in preparation and delivery
- General content of presentations made
- Contact person at service site
- Numbers of persons reached (estimated)
- Apparent outcome of the service

Reporting should be done quarterly, using the analysis techniques presented in Chapter 3 (the dropout prevention example).

Sample monitoring forms on the participant and the service delivery basis are in Appendix C.

Exhibit 4.1: *Possible Information to Include in a Participant Monitoring System*

Participant Data

- Age
- Sex
- Ethnicity
- Location of residence
- Socio-economic status
- Reason for involvement/referral source
- School or work status

Service Data

- Dates of first and last involvement
- Types and number of activities participated in
- Outcome of involvement

2. Consumer Satisfaction Surveys or variants thereof, such as interviews with program recipients, can serve to provide valuable data regarding how recipients view the program. Either quantitative or qualitative approaches can be used.[26] These surveys are very easy to conduct, and often yield positive and constructive feedback which can be readily used by programs. For example, responses to a satisfaction survey of students after the first semester in an experimental substance abuse prevention program indicated a large percentage felt not enough information on the effects of various drugs had been provided. The course was altered accordingly, and satisfaction responses at the end of second semester indicated that this was therefore no longer an issue.[19]

3. Implementation Analysis, based primarily on observation and interviews, can determine the extent to which a program model was actually put into place. It can also identify and document the process by which the program evolved. The evaluator, in such situations, serves as "program historian." Experiences at one site can often assist in replication elsewhere, as well as provide formative evaluation to the focal program. Since all programs are expected to change and evolve over time, implementation analysis should not be seen as a threat.

4. Participant Observation and Qualitative Interviewing are general research methods to gather data concerning any number of evaluation questions. The participant observer becomes immersed in the situation, experiencing it to some extent as a program participant would. **Field notes**—a sort of journal describing the setting, the participants, conversations,

and events, as well as the observer's reactions, comments, and analysis (separately delineated)—are prepared immediately after periods of observation, and serve as the primary data source. Time allocation for preparation of notes should be equal to the time spent observing or interviewing.

Observational work may be very general, attempting to analytically understand and describe a setting in its totality, or it may be very specific, focusing only on a particular type of event or problem. Analytic categories and understandings emerge with data collection, and may alter the course of future observation. While subjective experience serves as part of the data, the participant observer must attempt to maintain objectivity.

Observation should be systematic. This can be done by developing an observation plan in which the full range of events and situations relevant to the program are sampled, and similar types of data elicited. The goal of this style of research is to understand, rather than enumerate; thus, rarely will the qualitative analyst provide data on the frequency of occurrences.

Qualitative interviews seek in-depth data from the respondent's perspective. An interview guide (a short list of questions to be asked) may be used to make sure all important areas are covered, but an attempt is made to avoid influencing the respondent's answers, and to avoid structuring them by forcing them into one's own pre-conceived categories. (Most quantitative techniques structure everything in advance. Any response which does not fit the structure is treated as irrelevant or as a non-response.)

Qualitative techniques have the advantage of yielding rich, in-depth data from the program recipient's perspective. They also have the disadvantage of failing to yield enumerative data which is often called for. Qualitative research is labor intensive and requires a great deal of staff time to conduct. Qualitative data is also more easily distorted and misrepresented than quantitative data, as a result of partial perception, subjective interpretation, and selective recall by the evaluator. However, qualitative techniques are well suited to attempts to develop a comprehensive description and analysis of a program, particularly when used in conjunction with quantitative monitoring and outcome data.[23]

For example, participant observation and qualitative interviewing in the evaluation of one school-based prevention program revealed that the teachers who had been trained to conduct the activity ("Magic Circles") did not fully understand nor properly use the technique. The poor outcomes experienced by this program were attributed to improper implementation, as ascertained by qualitative process evaluation.[27]

Establishing a monitoring system from which regular data on program participants and services can be retrieved is crucial for any prevention program. Other process evaluation techniques, such as consumer satisfaction surveys, implementation analysis, participant observation, and qualitative interviewing are also invaluable in trying to provide a comprehensive description of what the program *really* is and to help understand program outcomes.

CHAPTER 5
OUTCOME EVALUATION

Outcome evaluation seeks to determine if a program achieved the desired change among its immediate recipients. The change, in the case of prevention programs, may actually be a *lack* of change where negative change would have occurred without the program. There is a great need in the prevention field for outcome studies, given that some prevention approaches may actually be harmful, others ineffective, and others helpful and effective. With sufficient rigorous outcome data, the field would advance by weeding out the ineffective and harmful techniques, as well as by strengthening and legitimating successful strategies.

Two questions must be answered in an outcome study:

1. Was there any change?
2. Was the change attributable to the program?

While the first question is often relatively easy to answer, the second question is fraught with complex methodological problems.

Experimental Control

Assuming the availability of adequate instrumentation, a simple pre- and post-test of recipients may be sufficient to answer the question of whether or not there was any direct change. However, there are many difficulties in attributing change to a program. These problems ("threats to validity") include the influence of **extraneous historical events; maturation of recipients** so that the "change" is actually a normally occurring one; **biased selection** into groups; **"mortality"** or dropout from the data collection process; and statistical **regression to the mean** (by which persons scoring to an extreme on a pre-test are likely to score less to the extreme on a post-test).[28]

Programs for the elderly, which seek to maintain a high quality of life, provide an example of these problems. Changes in funding for various necessary services may occur (historical event); participants get older and therefore become more likely to suffer from degenerative medical conditions (maturation); some participants may move away to live with relatives in a distant location and thus be unavailable for follow-up (mortality); and those most acutely in need may be selected for services (leading to regression to the mean). Any or all of these factors may intervene in such a manner that a simple one-group outcome study will yield inconclusive or uninterpretable results, or even make a successful program appear harmful.

Similarly, in substance abuse prevention programs for middle school students, maturation is a major problem. As students get older, they are more likely to experiment with alcohol and other drugs, whether or not they are exposed to prevention programming. Without adequate controls, such programs may appear to have harmful effects.

The problem of attributing change to a program rather than to extraneous factors such as maturation has led to the extensive use of experimental or quasi-experimental designs in evaluation. Experimental methods require random assignment to treatment and control groups, often impossible for practical or ethical reasons. Quasi-experimentation attempts to approximate experimental control:

- by using naturally occurring groups, similar to those receiving the program, as the control,
- by performing complex analysis of time-series data,
- by establishing a random tiebreaking procedure for clients equally eligible for a program when there are fewer slots than applicants available,
- or by using related techniques.

The key issue in designing outcome evaluations is that **comparisons** are made. Ideally, the groups which are compared were the same before exposure to the program, and, without the program, would have been expected to stay the same. This is assured by random assignment when true experimental designs are used. Quasi-experimentation only approximates this level of "control," and thus often yields ambiguous results.

Quasi-experimental approaches to outcome evaluation require complex statistical manipulations of data, and are generally expensive to implement. For the most part, experimental and quasi-experimental ap-

proaches belong to the "evaluation research" end of the continuum of evaluation, and often are not practical to incorporate in many local program evaluation efforts. School-based programs are an exception, in that one can easily select classrooms at random to serve as experimental and control groups.[29]

Uncontrolled Follow-up Studies

Although it is necessary to exercise some form of experimental or quasi-experimental control in order to attribute causality to the program, follow-up studies without control are also of value. Simple follow-up of participants, without control groups, can provide data on program success related to participant or service characteristics, and can give an indication of overall program performance. While scientifically of dubious value, such outcome studies are valuable to program evaluation (as opposed to evaluation research) efforts. They do provide evidence of change, even though it cannot be conclusively attributed to the program.

To be most useful, follow-up studies should a) carefully define the population to be included (e.g., all persons experiencing the program, or only those completing it), b) collect baseline data on client status before entry into the program, and c) take all possible steps to ensure a high rate of follow-up. Depending on resources available and the nature of the program, follow-up can include the use of mailed questionnaires (undesirable due to a relatively low return rate), telephone interviews, group administration of questionnaires, or in-person interviews. Collateral information (e.g., from parents or close friends) can also be obtained in follow-up, although its value has often been overstated.

To reduce the costs of follow-up outcome studies, random sampling of recipients is a useful strategy in programs which serve large number of participants. If true random selection for follow-up is achieved, it is safe to generalize to an entire treatment population based on the sample data.[30]

Immediate Outcomes

For some programs immediate change may be expected and relevant. In these cases, very simple pre- and post-test designs can be implemented, in which program participants answer a questionnaire at the first contact and at the end of the final session. The answers are compared to assess change. This approach is particularly useful in school-based prevention programs. The danger of such designs is that short-term change may have no relationship to lasting effects of the program. Changes in knowledge and attitudes, if measured, are likely to be evident in such pre- and post-test designs. However, we cannot be certain that these changes relate to longer term behavioral changes. A simple pre- and post-test design, however, has the advantage of low cost and high completion rates, since it is not necessary to locate the participant at follow-up. This design cannot be used in settings that have open-ended terminations.

Other prevention programs also lend themselves to immediate outcome evaluation. Dropout prevention can easily be assessed for successful outcome, as can genetics counseling and employment training. In each of these areas, there is a specific immediate outcome intended, and contact with participants usually lasts long enough to determine if the outcome objective has been attained. In such cases, outcome should be documented for each participant as part of an ongoing monitoring system (as was done in the dropout prevention example used to illustrate data analysis).

Community development programs, such as Positive Youth Development and Project NEWS, are less precise in their goals and objectives. Such programs usually have both intermediate and long-term objectives (e.g., to obtain youth participation in local planning groups may be an intermediate objective which, it is hoped will positively influence both the local youths and the community climate, in turn helping all youths). In these situations, the intermediate outcome may be easy to assess (i.e., do a substantial number of youth become meaningfully involved?), while the long-term outcome could require research far beyond the resources of most programs. With broad-aim programs such as these, special care must be taken in developing outcome objectives (see Chapter 3) to assure that expectations are set in advance against which to judge program performance.

The decision to undertake an outcome study must be made with full knowledge of likely methodological limitations and the necessary tradeoffs to be made between rigor and costs (including fiscal, staff cooperation, ethical, and program disruption costs). Expertise in statistical analysis beyond that present in most programs is required, particularly when quasi-experimental designs are used.

Given the limitations of design and measurement, along with the difficulty of implementing experimental designs in real world prevention programs, the results of local level outcome evaluation are likely to be suggestive but ambiguous and uncertain. Yet, it is also true that as data from many outcome studies accumulate we will gradually be able to sort out effective from ineffective program models, in spite of the methodological limitations of any one evaluation. Plausibility of results, rather than certainty, is the crucial issue. Thus, while we do not want to oversell the benefits of outcome evaluation for all programs, we wish to emphasize the ultimate importance of outcome studies to the field.

CHAPTER 6

CAUTIOUS EXPECTATIONS FROM EVALUATION

To some extent, the value of program evaluation has been oversold, both in the evaluation literature and (perhaps) even in this guide. Funders' expectations, in particular, have been overly ambitious given the available level of funding for prevention programs and the cost of doing rigorous evaluation research. In spite of the problems, evaluation is important, both to individual programs and to the prevention field as a whole.

While some requests for grant proposals call for comprehensive outcome, impact, and process evaluation, prevention practitioners should realize that this is unrealistic without extremely high funding levels in which the evaluation costs as much as, or more than, the program. And even if the high level of funds were forthcoming, we should be aware that the results of any one outcome or impact study are unlikely to be conclusive. Regardless of the level of funding, the apparent rigor of methodology, and sophistication of analysis, we are only able to build a case for the plausibility of our findings, and not their certainty.

We should also be aware that very few decisions are based on certain information. Rather, most decisions (in business, industry, government, and human services) are based on partial and limited information. Furthermore, decisions are not made on the basis of data alone (or even primarily). Rather, data is but one influencing factor brought to bear along with political pressures and personal interests. Data does, however, have some influence which can be exploited provided we have collected it. We must recognize that evaluation is, ultimately, a *political process*. Social programs, including prevention efforts, are developed and funded within a political arena, and are maintained or eliminated within a political structure of competition for scarce funds. Evaluative data, due to its usually mixed results, can provide ammunition for *all* sides in these contests. Often funders, practitioners, and decision-makers dismiss evaluative data unless it reinforces their preexisting opinions.

Evaluation cannot serve as a substitute for constituency organizing. Broad-based support, or even support from a few powerful figures, will probably be more effective in obtaining resources for prevention programming than positive findings from a rigorous evaluation. Given that prevention programs have not attained the level of institutionalized legitimacy that treatment and rehabilitation programs have, it may be that inequitable evaluation demands are being placed on prevention programs while the more institutionalized (and far more expensive) programs remain largely unevaluated. More evaluation of prevention will not solve these problems—they are political issues requiring political solutions.

Practitioners should also be aware that computers are not a magic solution to evaluation problems. One does not simply "run it through the computer," at least not yet. Rather, a great deal of knowledge, skill, and time are still required to handle quantitative data and to get the computer to do what you want with the data. The results must then also be interpreted to draw out their practical meaning.

Timing is also a crucial issue in prevention evaluation. Prevention programs, particularly those which are community oriented, take a great deal of time to implement and stabilize. Outcome evaluation should not be attempted before stabilization, and even then only if the program aims (goals, objectives) are specific and have a realistic chance of being attained. If the official goals and objectives (prepared for the funder) do not truly reflect the program's operative reality (as can be assessed in process evaluation), it is probably futile to attempt to measure "outcome" as defined by the official statement.

Hurriedly produced results may also be in error. If analyzed and presented to meet a deadline for funding or other crucial decisions, without time to thoroughly analyze results, major error may result. Process evaluation, describing and analyzing in-depth the nature of the program, collecting information on recipients, and examining implementation may be as much as should be done in many situations.

As generic prevention models continue to develop, it will be increasingly difficult to measure outcome, due to their imprecise multiplicity of goals and objectives. It will probably be necessary, if outcome data are collected at all, to examine intermediate outcomes such as decision-making skills and attitudes toward self and others, rather than ultimate outcomes such as "preventing alcohol abuse."

A further timing issue relates to the expected outcome of prevention. The ultimate preventive effect may not be seen for 10, 20 or 30 years after program exposure. It is, of course, not feasible for local programs to conduct the necessary follow-up studies over this time period to adequately judge their effectiveness. Large scale research, externally funded, is needed for this. It could be argued that large scale funding should be provided to rigorously test the efficacy of model demonstration programs. Once these models have been proven (or rejected and eliminated), effective models would be replicated elsewhere with a primary evaluation concern for faithful implementation of the model—a *process* concern.

Attempts to evaluate prevention programs will serve an educational function for those involved and lead to positive incremental program development and improvement. Practitioners will have a better understanding of what they are doing, a better overview of the full range of characteristics of program participants, and be better able to communicate to others about their programs. In this context, evaluation is best seen as evaluation *for*, rather than *of*, the program.

For most programs, monitoring and process evaluation is the most useful place to begin. Knowledge of program effectiveness (i.e., outcome) is important to the prevention field and must also be accumulated. We are convinced that, as data are brought to bear, it will become increasingly obvious that prevention works.

REFERENCE NOTES

1. Wisconsin Department of Health and Social Services, "Prevention Statement," (Madison, WI: DHSS, 1982) p.1.

2. Wisconsin Prevention Network, "What is Prevention?" (Madison, WI: WPN, 1982) p.1.

3. The emphasis on usefulness is stressed in evaluation resources such as: C. Clifford Attkisson, W.A. Hargreaves, M.J. Horowitz and J.E. Sorenson, *Evaluation of Human Service Programs* (New York: Academic Press, 1982); and Michael Q. Patton, *Utilization-Focused Evaluation* (Beverly Hills: Sage, 1978). The reader interested in increasing the usefulness of evaluation is urged to consult these sources.

4. This distinction was first proposed by Michael Scriven, "The methodology of evaluation," pp. 39-83 in R. Tyler et. al. (Editors), *Perspectives of Curriculum Evaluation* (Chicago: Rand McNally, 1967).

5. This classification was developed by the National Institute on Drug Abuse (NIDA), and has been promulgated in John F. French and Nancy J. Kaufman (Editors), *Handbook for Prevention Evaluation* (Rockville, MD: National Institute on Drug Abuse, 1981) DHSS Publication No. (ADM) 81-1145. While the term "process evaluation" is fairly standard in usage, the evaluation literature often uses the terms "outcome," "impact" and "effectiveness" interchangeably.

6. John W. Meyer and Brian Rowen, "Institutionalized organizations: Formal structure as myth and ceremony." *American Journal of Sociology* 83(2):340-363; 1977.

7. Irving Piliavin and Thomas MacDonald, "On the fruits of evaluative research for the social services," pp. 337-344 in S. Slavin (Editor), *Social Administration* (New York: Haworth Press, 1978).

8. Joan McCord, "Consideration of some effects of a counseling program," pp. 394-405 in S.E. Martin et al. (Editors), *New Directions in the Rehabilitation of Criminal Offenders* (Washington D.C.: National Academy Press, 1981).

9. One example of a comprehensive outcome monitoring system is that used by Lutheran Social Services of Wisconsin reported by William R. Benedict and Dennis A. Bruce, "A social systems programme evaluation and management approach for human service organizations" pp. 2-9 in B. Willer et al. (Editors), *Information and Feedback for Evaluation* (Toronto, Canada: York University Communications Dept., 1975).

10. Robert A. Bell, T.D. Nguyen, G.J. Warheit, and J.M. Buhl, "Service utilization, social indicators, and citizen survey approaches to human service needs assessment," pp. 253-300 in C. Attkisson et al., (Editors), *Evaluation of Human Service Programs* (New York: Academic Press, 1978).

11. Carol H. Weiss, *Evaluation Research* (Englewood Cliffs: Prentice-Hall, 1972).

12. Judith Blanton and Sam Alley, "Machiavellian discourses on research and evaluation," *Evaluation* 2(1):18-19; 1974. Also, see Carol H. Weiss, "Evaluation research in the political context," pp. 13-26 in E. Struening and M. Guttentag (Editors), *Handbook of Evaluation Research, Vol. 1* (Beverly Hills: Sage, 1975).

13. Michael W. Apple, "The process and ideology of valuing in educational settings" in M.W. Apple et al., (Editors), *Educational Evaluation: Analysis and Responsibility* (Berkeley: McCutchan, 1974).

14. David Mechanic, "Evaluation in alcohol, drug abuse and mental health programs: Problems and prospects," pp. 3-25 in J. Zusman and C.R. Wurster (Editors), *Program Evaluation* (Lexington, MA: D.C. Heath, 1975).

15. These steps have been adapted from Michael Patton's *Utilization-Focused Evaluation* (See footnote 3.)

16. Maximizing usefulness is also stressed by Lee J. Cronbach and Associates in their book *Toward Reform of Program Evaluation* (San Francisco: Jossey-Bass, 1980).

17. This technique was developed by Michael Klitzner.

18. These steps are a summary of a booklet by Joel Cantor, Nancy Kaufman, and Michael Klitzner titled *Four Steps to Better Objectives*, available through Wisconsin Clearinghouse, 1954 E. Washington Ave., Madison, WI 53704, $2.50 postpaid. The reader is urged to obtain a copy of this booklet for more detail on this process.

19. D. Paul Moberg, Douglas Piper, and Gjerald Thompson, *Cross Age Drug Abuse Prevention Program: Final Report to Wisconsin's Dept. of Public Instruction* (Monona, WI: Monona Grove Public Schools, 1983).

20. Irwin Deutscher and Margaret Gold, "Traditions and rules as obstructions to useful program evaluation," *Studies on Symbolic Interaction* 2:107-140; 1979.

21. Several evaluation primers are available which discuss relatively simple analysis techniques. See especially Arlene Fink and Jacqueline Kosecoff, *An Evaluation Primer* (Washington D.C., Capitol Publications, 1978). Also useful to practitioners is a series of basic manuals on program evaluation by Lynn L. Morris, Carol T. Fitz-Gibbon, and Marlene E. Henerson, collectively referred to as the *Program Evaluation Kit* (Beverly Hills, CA: Sage, 1978). Also useful is the book by Michael J. Austin et al., *Evaluating your Agency's Program* (Beverly Hills, CA: Sage, 1983).

22. Although most prevention practitioners will need external assistance for statistical analysis, several resources are fairly accessible and easy to use by those with minimal statistical background. Beyond the resources referenced in Note 21, consult Stephen Isaac and William B. Michael, *Handbook in Research and Evaluation* (San Diego: Robert R. Knapp, 1971); Fred N. Kerlinger, *Foundations of Behavioral Research* (New York: Holt, Rinehart and Winston, 1973); and Hubert M. Blalock *Social Statistics* (New York: McGraw-Hill, 1979). The *Minitab* computer package is also easy to use for data analysis, and the program's manual provides relatively simple formulas and examples which can assist in analysis with or without a computer. See Thomas A. Ryan, Brian L. Joiner, and Barbara F. Ryan, *Minitab Student Handbook* (North Scituate, MA: Wadsworth, 1976).

23. For help in conducting qualitative analysis, see Michael Q. Patton, *Qualitative Evaluation Methods* (Beverly Hills, CA: Sage, 1980); and John Lofland *Analyzing Social Settings* (Belmont, CA: Wadsworth, 1971).

24. For findings on the use of evaluation, see Gary Cox, "Managerial style: Implications for the utilization of program evaluation information," *Evaluation Quarterly* 1:499-507, 1977; also, Carol H. Weiss, *Using Social Research in Public Policymaking* (Lexington, MA: D.C. Heath, 1977).

25. The complete evaluation of Project NEWS is available in a report by Lori Hahn and D. Paul Moberg, *Project NEWS: Final Program Evaluation Report* (Madison, WI: United Neighborhood Centers, 1983).

26. See Barry A. Tanner, "A multi-dimensional client satisfaction instrument," *Evaluation and Program Planning* 5:161-167, 1982; D.L. Larsen et al., "Assessment of client/patient satisfaction: Development of a general scale," *Evaluation and Program Planning* 2:197-207, 1979; and D. Paul Moberg, "Consumer evaluation of an alcoholism treatment program," pp. 525-534 in F.A. Seixas (Editor), *Currents in Alcoholism, Vol. 2* (New York: Grune and Stratton, 1977).

27. Joel M. Moskowitz, Eric Schaps, and Janet H. Malvin, "Process and outcome evaluation in primary prevention: The Magic Circle Program," *Evaluation Review* 6(6):775-788, 1982.

28. See Donald T. Campbell and Julian C. Stanley, *Experimental and Quasi-Experimental Designs for Research* (Chicago: Rand McNally, 1963).

29. For more information on experimental and quasi-experimental design, see Campbell (note 28); Thomas D. Cook and Donald T. Campbell, *Quasi-Experimentation* (Chicago: Rand McNally, 1979); and Henry W. Reicken, "Principal components of the evaluation process," *Professional Psychology* 8(4):392-410, 1977.

30. Many useful suggestions for conducting follow-up studies are provided by Lloyd Johnston et al., *Conducting Follow-up Research on Drug Treatment Programs* (Washington, D.C.: U.S. Government Printing Office, 1977). DHEW #(ADM) 78-487.

APPENDIX A
GLOSSARY OF TERMS

Archival Data: data collected by someone other than the evaluator for purposes other than the evaluation. Clinical records, school records, police records, and census data are examples.

Descriptive Statistics: statistical analysis which presents data describing the characteristics of samples or populations. Frequency distributions and measures of central tendency (mean, median, mode) are the most commonly used descriptive statistics.

Design: refers to the manner in which people or events are grouped and variables manipulated in order to answer the questions posed in an evaluation. (See experimental design, below.)

Efficacy: the level of effectiveness of a program for a defined target population under ideal conditions of program implementation.

Efficiency: the relative cost of program strategy is less in comparison to other strategies for achieving the same level of results.

Effectiveness: the level of benefit of a particular program strategy for a defined target population under average or typical conditions of program implementation (compare to efficacy).

Experimental Design: an evaluation design in which potential program participants are randomly assigned into experimental and control groups. Those in the control group either are provided no services or an alternative service to the experimental program.

Formative Evaluation: an evaluation oriented to providing ongoing feedback to program decision-makers to be used to improve the program. (Compare to summative evaluation.)

Impact Evaluation: an evaluation which focuses on the generalized effects of a program in a community, school, or other larger target population which goes beyond the immediate, direct program participants. (This terminology is not universally used in the evaluation field.)

Inferential Statistics: statistical analysis which generalizes to a population based on known characteristics of samples from the population. Inferential statistics are based on mathematical probability theory, and include such techniques as t-tests, chi-square, analysis of variance, and multiple regression analysis.

Instrumentation: the measuring devices used in an evaluation, which may involve questionnaires, interview schedules, standardized tests, or human observers.

Monitoring: ongoing assessment of the extent to which a program 1) conforms to its design, plans and regulations; 2) is reaching its specified target population; 3) is delivering the intended level, volume, and nature of services, and/or is meeting the need for its services.

Monitoring System: ongoing data collection and reporting mechanism to obtain information needed for monitoring purposes.

Needs Assessment: a systematic approach to determining the level of "need" in a particular geographic area or population for a proposed program or service. Techniques include analysis of social indicators, social area surveys (of citizens and/or service providers), and community group approaches (forums, nominal groups, delphi techniques, etc.).

Outcome Evaluation: assessment of the extent to which a program attains its objectives related to short- and long-term change in program participants' behavior, attitudes, knowledge, or levels of problems and dysfunctions. For community change programs, assessment of the extent to which the intended change occurs. (This terminology is *not* universal in the evaluation field.)

Operationalization: the transformation of the concepts which make up a program's theory or goals into measurable indicators.

Participant Observation: an approach to data collection in which the researcher becomes a part of the program while systematically observing and analyzing the ongoing activity. The participant observer is not necessarily a provider or recipient of the services. Field notes (the primary data record) are prepared during or after the observation.

Process Evaluation: description, analysis, and assessment of program inputs and activities. Process evaluation considers the *means* by which a program seeks to attain its outcome objectives. Process evaluation can address need for a program, the actual implementation of program theory and plans, staff and participant interaction, and actual delivery of services. Monitoring is one part of process evaluation.

Program: is used generically to refer to an organized effort to effect some human problem or conditions. Programs can range from small local level efforts staffed by one or two paid staff or volunteers at a single site (e.g., telephone hot lines) to complex, multi-site (often nationwide) efforts, such as AFDC, Food Stamps, CMHC's, or SSI.

Program Evaluation: the systematic collection, analysis, and interpretation of data for the purpose of determining the value of a social policy or program, to be used in decisionmaking about the policy or program.

Qualitative Data: systematically collected information which uses a human observer as the instrument. Qualitative data are verbal, narrative, or pictorial and do not readily lend themselves to quantification. Qualitative data are usually collected by participant observation and/or intensive interviewing.

Quantitative Data: systematically collected information which yields numerical representation. Quantitative data are usually collected by questionnaire, structured interview forms, tests, and/or simply by counting the frequency or intensity of particular events.

Quasi-Experimental Design: research and evaluation designs which attempt to approximate experimental designs when random assignment to groups is not feasible. The most common quasi-experimental designs are the non-equivalent comparison group and time series analysis.

Relevance: the usefulness of an instrument within the programmatic context, including consistency with program objectives and program definitions, appropriateness to program target population, usefulness in relationship to the community's culture, and believability to key decision-makers.

Reliability: the extent to which a measure is internally consistent and/or stable (i.e., yields the same results over repeat administrations assuming all relevant factors remain the same).

Summative Evaluation: an evaluation which assumes a stable and consistent program and yields after-the-fact judgements of program effectiveness, usually without ongoing feedback of results.

Unobtrusive Measures: data obtained in a manner in which there is no burden put on those who are the subjects of the evaluation. Hidden observation, analysis of physical traces, and use of archival records are examples.

Validity: the extent to which an observed or measured situation reflects the "true" situation, or, alternatively, the extent to which a particular measure adequately reflects the concept it is intended to measure.

These materials developed by D. Paul Moberg and Michael D. Klitzner. Acknowledgement of source appreciated.

APPENDIX B
DATA COLLECTION INSTRUMENT CHECKLIST*

I. **General:**

A. Are all relevant issues addressed?
B. Is reading level or language appropriate to age, cultural identity, educational level?
C. Are jargon and abbreviations avoided?

II. **Questions:**

A. Are all questions answerable by all respondents or in all cases?
B. Does each question address a single issue?
C. Are there biases in the way questions are phrased?
D. Are there both positively and negatively phrased questions?

III. **Response Categories:**

A. Unambiguous
B. Mutually exclusive
C. Exhaustive (include "other" when in doubt)
D. Too few or too many
E. Match the question

IV. **Format of Instrument:**

A. Is it clear how to respond to each item?
B. Is it clear when multiple responses (i.e., check all that apply) are allowed?
C. Is there enough room to respond?
D. Has consideration been given to ease of coding and retrieval?
E. Are responses precoded?
F. Does the sequence of items flow logically?

Developed by Michael Klitzner, Paul Moberg and Katherine Stewart. Acknowledgement of source appreciated.

APPENDIX C
SAMPLE PROGRAM MONITORING FORMS*

These forms were developed by a chemical dependency prevention and intervention center* which provides information and referral services to individuals as well as delivering education, training, and consultation services.

The first form is used for all information, referral, and intervention cases, monitoring program activities on the level of individual participants.

The second form is used to monitor service delivery activities which do not involve individual participants, such as inservice training, program consultation, and media campaigns.

Both forms are set up for computer processing.

Used with permission from the Prevention and Intervention Center for Alcohol and other Drug Abuse (PICADA), 17 N. Webster St., Madison, WI 53703.

PICADA EDUCATION/TRAINING CONTACT SHEET

ID Batch No.

Organization name	Address	Phone
Contact person(s)	Position	Phone

SERVICE RECORD

Date	Amt. of time	Time of day	staff codes			type

Amount of Time:
in hours and tenths

Time of Day:
1. regular office hours
2. after office hours

Type of Contact:
1 = consultation (phone)
2 = educational program
3 = training program
4 = networking
5 = planning/preparation

PREVIOUS CONTACT WITH PICADA
(check all that apply)

a ☐ School/agency/organization inservice training
b ☐ MMSD Learning Exchange
c ☐ Classroom programs
d ☐ Community Education programs
e ☐ Consultation
f ☐ Individual contact
g ☐ Media/public information
h ☐ Networking/Task forces
i ☐ EAP
j ☐ Other
k ☐ None—If none, referral source(s)

LOCATION OF PROGRAM

TYPE OF PROGRAM (check only one)

1. School–Youth Programs
2. School–Staff Programs
3. School–Consults
4. School–Networking
5. Youth and Family–Community Organization
6. Youth and Family–Education Programs
7. Youth and Family–Consults
8. Youth and Family–Networking
9. Youth Service Provider–Training
10. Adults–Education Programs
11. Adult–Consults
12. Adult–Networking
13. Adult Service Provider Training
14. EAP–Contracts
15. EAP–Consults
16. Public Info/Media

PROGRAM CONTENT (check only one)

1. Prevention
2. Intervention
3. Prev. & Inter.
4. AOD Information
5. What is PICADA

NUMBERS SERVED

Under 18

Adults

STAFF RESPONSIBLE FOR PROGRAM

Staff codes

☐ Were not able to meet complete request? If not, explain in program notes section.

Special Target Population
(if appropriate)

a ☐ Women
b ☐ Gay
c ☐ Cultural Minority
d ☐ Economic Minority
e ☐ Other (specify)

Program Request Notes

Date(s)

Time

Length of program

Location

Audience type and estimated number

Related training

Content

PICADA CONTACT SHEET

NAME ADDRESS PHONE

ID CODE SEQ

MATERIAL SENT BY / DATE

SERVICE TYPE
1 Information and Referral
2 Intervention
3 Agency Consultation
4 Family Education
5 Phone Assessments

CASE MGR

Referral Source

PRIMARY NAME

SECONDARY NAME

SEX
1 Female
2 Male

ETHNICITY
1 White
2 Black
3 Latin
4 Native American
5 Asian
6 Other
7 Unknown

AGE
If unknown, round to the nearest 5 except minors = 15

PRESENTING PROBLEM INVOLVES (check one)
1 Drugs
2 Alcohol
3 Poly-drug
4 Non-AODA problem
5 Multiple problems
6 Undetermined
7 Other (specify)

CLIENT'S RELATIONSHIP TO PROBLEM (check one)
1 Self
2 Parent
3 Spouse/Partner
4 Sibling
5 Child
6 Other Relative
7 Friend/Acquaintance
8 EAP Staff/Co-worker
9 University Staff
10 School Staff
11 Agency Staff

TYPES OF DRUGS INVOLVED (Check all that apply)

- ☐ a Alcohol
- ☐ b Barbiturates
- ☐ c Other Depressants
- ☐ d Narcotics
- ☐ e Amphetamines
- ☐ f Other Stimulants
- ☐ g Cocaine
- ☐ h Marijuana
- ☐ i Psychedelics
- ☐ j Inhalants
- ☐ k Over-the-counter
- ☐ l Other prescription
- ☐ m General drug info
- ☐ n Other
- ☐ o Unknown drug
- ☐ p Not applicable

CLIENT OR PROBLEM INVOLVES EAP School University

ID Code _____

SERVICE RECORD

Date	amt. of time	time of day	staff codes		type

Amount of time: in hours and tenths
Time of Day: 1 Regular office hours
2 After office hours

Type 1 = phone 2 = walk in 3 = correspondence
4 = out of office 5 = appointment 6 = follow-up
7 = F. Ed. lecture 8 = F. Ed. Assmt. 9 = No Show

SERVICES PROVIDED (Check all that apply)

- ☐ a Info on PICADA
- ☐ b General AOD info
- ☐ c Referral
- ☐ d Drug ID
- ☐ e Library
- ☐ f AA/Alanon Sched.
- ☐ g Handouts
- ☐ h Consultation
- ☐ i Assessment
- ☐ j Other

REFERRALS TO: CC (Completion Code) 1 = Referral followed; 2 = Referral Not Followed; 3 = Unknown; 4 = Not Available

CODE	CC	AGENCY NAME

NARRATIVE

BEST COPY AVAILABLE

PICADA CONTACT SHEET
PAGE 2

NAME	APPT. DATE	APPT. TIME	APPT. WITH:

ID CODE | SEQ. | BIRTH DATE

GEOGRAPHICAL LOCATION

INTERVENTION PLAN

1 NO PLAN NEEDED
2 NO PLAN AGREED UPON
3 PLAN AGREED UPON

IF PLAN, IT WAS
1 FOLLOWED
2 PARTIALLY FOLLOWED
3 NOT FOLLOWED
4 OUTCOME UNKNOWN

EMPLOYMENT STATUS

1 Employed
2 Unemployed
3 House Manager
4 Student
5 Retired
6 Other _____
7 Unknown

MARITAL STATUS

1 Single
2 Married
3 Living with partner
4 Separated
5 Divorced
6 Widowed
7 Unknown

FINANCIAL RESPONSIBILITY (If referral check one)

1 Self
2 Insurance
3 51.42
4 Other _____

PRIOR SERVICE NETWORK INVOLVEMENT
(Check if prior involvement)

- ☐ a AODA-NA/AA/Alanon
- ☐ b AODA-Outpatient
- ☐ c AODA-Inpatient
- ☐ d Mental Health Outpatient
- ☐ e Mental Health Inpatient
- ☐ f Other Social Services
- ☐ g Legal System

FINDINGS OF ASSESSMENT
AOD use (check only one)

1 Dependent
2 Abusing
3 Experimental use
4 Non-Problem Use
5 Non-User
6 Recovering/non-using dependent
7 Uncertain/cannot determine

PROBLEM AREAS (check all that apply)

- ☐ a Family
- ☐ b Marital
- ☐ c Sexual
- ☐ d Child Abuse
- ☐ e Physical/Sexual Abuse _____
- ☐ f Mental Health
- ☐ g Work Adjustment
- ☐ h Financial
- ☐ i Legal
- ☐ j School
- ☐ k Physical Handicap
- ☐ p Physical Health
- ☐ m Other _____

TOTAL NUMBER OF CONTACTS WITH SIGNIFICANT OTHERS OR COLLATERALS (Specify 0-8; 9 = 9 or more)

- ☐ Referring Agent
- ☐ Agency Referred To
- ☐ Family Members
- ☐ Friends
- ☐ Employer
- ☐ School
- ☐ Others (specify) _____

NARRATIVE